

We claim:

1. A method for tracking video objects in video frames, the method comprising:

5 performing spatial segmentation on a video frame to identify regions of pixels with homogenous intensity values;

performing motion estimation between each of the regions in the video frame and a previous video frame;

using the motion estimation for each region to warp pixel locations in each region to locations in the previous frame;

10 determining whether the warped pixel locations are within a boundary of a segmented video object in the previous frame to identify a set of the regions that are likely to be part of the video object; and

forming a boundary of the video object in the video frame as a combination of each of the regions in the video frame that are in the set.

15

2. The method of claim 1 further including:

repeating the steps of claim 1 for subsequent frames using the boundary of the video object as a reference boundary for the next frame.

20

3. The method of claim 1 further including:

filtering the video frame to remove noise from the video frame before performing the spatial segmentation.

25

4. The method of claim 1 wherein each of the regions are connected group of pixels, and wherein each region is determined to be homogenous by ensuring that the difference in intensity values between a pixel location with a maximum intensity value in the region and another pixel location with a minimum intensity value region is below a threshold.

5. The method of claim 4 wherein the segmentation is sequential region growing method comprising:

starting with a first pixel location in the video image frame, growing a first region of connected pixels around the first pixel by adding pixels to the region such

5 that the homogeneity criteria is satisfied;

when no boundary pixels satisfy the homogeneity criteria, repeating the growing step with a pixel location outside the first region; and

continuing the growing step until each of the pixels in a frame is identified as being part of a homogenous region.

10

6. The method of claim 1 wherein the motion estimation comprises:

for each region identified through spatial segmentation in the video frame, performing a region based motion estimation including matching only pixels within the region with pixels in the previous frame to find a corresponding location for each

15 of the pixels in the previous frame; and

applying a motion model to approximate motion of the pixels in the region to the corresponding locations in the previous frame.

7. The method of claim 6 wherein the motion model is used to find a motion vector for each region that minimizes prediction error between warped pixel values from the video frame and pixel values at the corresponding pixel locations in the previous video frame.

20

8. The method of claim 1 wherein the determining step includes:

25 finding the number of warped pixels that are inside the boundary of the segmented video object of the previous frame;

when a majority of the warped pixels lie inside the boundary of the segmented video object, classifying the region as being part of the video object in the video frame.

30

9. A computer readable medium having instructions for performing the steps of claim 1.

10. A computer readable medium having instructions for tracking semantic
5 objects in a vector image sequence of image frames, the medium comprising:

a spatial segmentation module for segmenting a vector image frame in the image sequence into regions, each region comprising connected groups of image points having image values that satisfy a homogeneity criterion;

a motion estimator module for estimating the motion between each of the
10 regions in the input image frame and a reference frame and for determining a motion parameter that approximates the motion of each region between the image frame and the target frame; and

a region classifier for applying the motion parameter of each region to the region to compute a predicted region in the reference frame, for evaluating whether a
15 boundary of each predicted region falls at least partially within a boundary of a semantic object of the reference frame, and classifying each region as being part of semantic object in the reference frame based on the extent to which the predicted region falls within the boundary of a semantic object boundary of the reference frame;

wherein a boundary of a semantic object in the image frame is formed from
20 each region classified as being part of a corresponding semantic object in the reference frame.

11. The method of claim 10 wherein the homogeneity criteria of the spatial segmentation module comprises a maximum difference value between a first image
25 point in a connected group of pixels with a maximum image value and a second image point in the connected group with a minimum image value, and wherein the segmentation module selectively adds neighboring image points to the connected region to create a new connected region so long as the new connected region satisfies the homogeneity criteria.

12. The method of claim 10 wherein the motion parameter of each region is a motion vector, which, when used to project each image point in a region into the target frame, minimizes a sum of differences between image values of the projected points and image values at corresponding image points in the target frame.

5

13. The method of claim 10 wherein:

the target frame includes two or more semantic objects, each object occupying a non-overlapping area of the target frame,

the region classifier identifies for each predicted region, a semantic object in
10 the target frame having a maximum number of overlapping image points of the predicted region,

the classifier classifies each region as being associated with a semantic video object in the target frame having the maximum number of overlapping image points, and

15 the classifier computes boundaries of each semantic object in the image frame as a combination of regions classified as being associated with the corresponding semantic object in the target frame.

14. The medium of claim 13 further including:

20 a majority operator for defining a structure of points around each image point in the image frame, for determining a semantic object in the image frame that has a maximal overlapped area of the structure, and for assigning a value of the semantic video object to the image point.

25 15. A method for tracking semantic objects in vector image sequences, the method comprising:

performing spatial segmentation on an image frame to identify regions of discrete image points with homogenous image values;

performing motion estimation between each of the regions in the image frame
30 and a target image frame in which a boundary of a semantic object is known;

using the motion estimation for each region to warp the image points in each region to locations in the target frame;

determining whether the warped pixel locations of each region are within a boundary of a semantic object in the target frame and when at least a threshold amount
5 of the region overlaps a semantic object in the target frame, classifying the region as originating from the semantic object in the target frame; and

forming a boundary of the semantic object in the image frame as a combination of each of the regions in the image frame that are classified as originating from the semantic object of the target frame.

10

16. The method of claim 15 further including:

repeating the steps of claim 15 for subsequent frames using computed boundaries of semantic objects of a previous frame to classify regions segmented in a current frame as originating from one of the semantic objects of the previous frame.

15

17. The method of claim 15 wherein each of the regions are connected group of image points, and wherein each region is determined to be homogenous by only adding neighboring image points to the region where the difference in intensity values between a maximum and minimum image value in the region after adding each
20 neighboring image point is below a threshold.

18. The method of claim 15 wherein:

the target frame is the previous frame of the current frame,
each region segmented from the current frame is classified as originating
25 from exactly one semantic object previously computed for the previous frame using the steps of claim 15,

boundaries for semantic objects in the current frame are computed by combining boundaries of regions classified as originating from the same semantic object in the previous frame, and

the steps of claim 15 are repeated for successive frames in the vector image sequence.

19. A computer readable medium having instruction for performing the steps
5 of claim 15.

20. A method for tracking semantic objects in vector image sequences, the method comprising:

performing spatial segmentation on an image frame to identify regions of
10 discrete image points with homogenous image values, where each of the regions are connected group of image points, and where each region is determined to be homogenous by only adding neighboring image points to the region where the difference in intensity values between a maximum and minimum image value in the region after adding each neighboring image point is below a threshold;

15 performing region based motion estimation between each of the regions in the image frame and an immediate previous image frame in the vector image sequence;
using the motion estimated for each region to warp the image points in each region to locations in the immediate previous frame;

determining whether the warped pixel locations of each region are within a
20 boundary of a semantic object in the target frame and when at least a threshold amount of the region overlaps a semantic object in the target frame, classifying the region as originating from the semantic object in the target frame; and

forming a boundary for each semantic object in the image frame as a
combination of each of the regions in the image frame that are classified as originating
25 from the semantic object of the immediate previous frame.